Tennessee Higher Education Commission

Supplementary Task: Creating Models for Data

This task is designed to be a collection of data sets that can be used as classroom examples or as tasks for student work. Not all data sets provide clearly-determined solutions.

Culminating Project: Creating Models for Data

Data collected for various situations is presented below. For each set of data, students should be asked to:

a) Determine what type of model is most appropriate for the data. Explain what models were considered and why one model was chosen over the others. Explain what each of the variables represents.
b) Present graphical evidence that the model is valid for the data presented.
c) Explain any limitations of the model.
d) Create several questions that can be answered using the model. Answer the questions that are created.

Data Set 1: Diameter of Sand Granules vs. Slope of the Beach

For naturally occurring ocean beaches, the data below were collected. Find a function to model the relationship between the median diameter of granules of sand (in millimeters) and the gradient of the slope of the beach (in degrees).

Diameter	0.17	0.19	0.22	0.235	0.235	0.3	0.35	0.42	0.85
Gradient	0.63	0.7	0.82	0.88	1.15	1.5	4.4	7.3	11.3

Resource: *Physical Geography*, by A. M. King, Oxford Press, England

Data Set 2: Cricket Chirps

The following data relate the number of chirps per second of the striped ground cricket and the temperature in degrees Fahrenheit. Find a function to model the relationship between these two quantities.

Chirps/sec	20.0	16.0	19.8	18.4	17.1	15.5	14.7	17.1	15.4	16.2	15.0	17.2	16.0	17.0
Temperature	88.6	71.6	93.3	84.3	80.6	75.2	69.7	82	69.4	83.3	79.6	82.6	80.6	83.5

Resource: *The Song of Insects*, by Dr. G. W. Pierce, Harvard College Press

Data Set 3: Ground Water Survey

The following data were collected from a random sample of wells in Northwest Texas. The data relate the bicarbonate (in parts per million) of the well water with the pH of the well water. Use the data to develop a model for the relationship between these two quantities.

Bicarbonate	157	174	175	188	171	143	217	190	142	190	215	199	262	105
pH	7.6	7.1	8.2	7.5	7.4	7.8	7.3	8.0	7.1	7.5	8.1	7.0	7.3	7.8

Resource: Union Carbide Technical Report K/UR-1

Data Set 4: Prehistoric Pueblos

The data below relate the estimated year of initial occupation with the estimated year of the end of occupation of prehistoric pueblos in a random sample of such pueblos in Utah, Arizona, and Nevada. Use the data to develop a model for the relationship between these two quantities.

Initial Year	900	700	1125	750	1250	1250	1175	1225	1180	1080	1080	1075	1090	1225	1200	1325
End Year	1250	1300	1175	1250	1300	1280	1225	1275	1250	1150	1275	1250	1135	1275	1285	1400

Resource: *Prehistoric Pueblo World*, by A. Adler, University of Arizona Press

Data Set 5: The Size of Alligators

Many wildlife populations are monitored by taking aerial photographs. For example, the length of an alligator can be estimated quite accurately from an aerial photograph. However, the alligator's weight is much more difficult to determine. The data below show the length (in inches) and weight (in pounds) of alligators captured in central Florida. Use the data to develop a model for the relationship between the length and weight of alligators in central Florida.

Length	58	61	63	68	69	72	72	74	74	76	78	82	85	86	86	86	88	89	90	90	94	94	114	128	147
Weight	28	44	33	39	36	38	61	54	51	42	57	80	84	83	80	90	70	84	106	102	110	130	197	366	640

Resource: Exploring Data website, http://curriculum.qed.qld.gov.au/kla/eda, Education Queensland, 1997

Data Set 6: Sunflower Height

The data below relates the height of single sunflower plant (in cm) to the number of days the plant has been growing. Use the data to develop a model for the relationship between the height and the number of days the plant has been growing.

Day	0	7	14	21	28	35	42	49	56	63	70	77	84
Height	0.00	17.93	36.36	67.76	98.10	131.00	169.50	205.50	228.30	247.10	250.50	253.80	254.50

Resource: Reed, H.S. and Holland, R.H. (1919), Growth of sunflower seeds; Proceedings of the National Academy of Sciences, volume 5, p. 140.

Data Set 7: Moving River Particles

Rivers and streams carry small solid particles of rock and mineral downhill, either suspended in the water column or bounced, rolled, or slid along the river bed. The data below show the speed (in m/sec) necessary to carry particles in suspension as it relates to the diameter of the particle (in mm). Use the data to develop a model for the relationship between the size of the particle in suspension and the speed necessary to move the particle.

Diameter	0.2	1.3	5	11	20	45	80	180
Speed	0.10	0.25	0.50	0.75	1.00	1.50	2.50	3.50

Resource: Nielsen, A. (1950) *Oikos,* 2, 176-96 as reported in Ecology for Environmental Sciences, Anderson J.M.

Data Set 8: World Records for Men's High Jump

The data below show the world record for the men's high jump (in meters) and the year in which the world record was set (between 1912 and 1993). Use the data to create a model for the relationship between the world record for the men's high jump and the year in which the record was set.

Year	1912	1914	1924	1933	1934	1936	1937	1941	1953	1956	1957	1960	1960
Record	2.00	2.01	2.03	2.04	2.06	2.07	2.09	2.11	2.12	2.15	2.16	2.17	2.18

(table continued)

Year	1960	1961	1961	1961	1962	1962	1963	1971	1973	1976	1976	1977	1978
Record	2.22	2.23	2.24	2.25	2.26	2.27	2.28	2.29	2.30	2.31	2.32	2.33	2.34

(table continued)

Year	1980	1980	1983	1983	1984	1985	1985	1987	1989	1993
Record	2.35	2.36	2.37	2.38	2.39	2.40	2.41	2.42	2.44	2.45

Data Set 9: World Records for Women's High Jump

The data below show the world record for the women's high jump (in meters) and the year in which the world record was set (between 1932 and 1987). Use the data to create a model for the relationship between the world record for the women's high jump and the year in which the record was set.

Year	1932	1939	1943	1951	1954	1956	1956	1956	1957	1958	1958	1958	1958
Record	1.65	1.66	1.71	1.72	1.73	1.74	1.75	1.76	1.77	1.78	1.80	1.81	1.82

(table continued)

Year	1958	1958	1960	1960	1961	1961	1961	1971	1972	1974	1976	1977	1977
Record	1.83	1.84	1.85	1.86	1.88	1.90	1.91	1.92	1.94	1.95	1.96	1.97	2.00

(table continued)

Year	1978	1982	1983	1983	1984	1984	1986	1987
Record	2.01	2.02	2.03	2.04	2.05	2.07	2.08	2.09

Data Set 10: Space Shuttle Ascent Altitude

On July 4, 2006, the Space Shuttle Discovery launched from Kennedy Space Center on mission STS-121 to begin a rendezvous with the International Space Station. The data below show the altitude (in feet) of Discovery every 10 seconds from liftoff to the separation of the solid rocket boosters. Use the data to model the altitude of Discovery at time t.

Time	0	10	20	30	40	50	60	70	80	90	100	110	120
Altitude	7	938	4160	9872	17635	26969	37746	50548	66033	83966	103911	125512	147411

Resource: www.nasa.gov

Data Set 11: Space Shuttle Ascent Total Mass

On July 4, 2006, the Space Shuttle Discovery launched from Kennedy Space Center on mission STS-121 to begin a rendezvous with the International Space Station. The data below show the total mass (in kg) of Discovery every 10 seconds from liftoff to the separation of the solid rocket boosters. Use the data to model the total mass of Discovery at time t.

Time	0	10	20	30	40	50	60	70	80	90	100
Total Mass	2,051,113	1,935,155	1,799,290	1,681,120	1,567,611	1,475,282	1,376,301	1,277,921	1,177,704	1,075,683	991,872

(table continued)

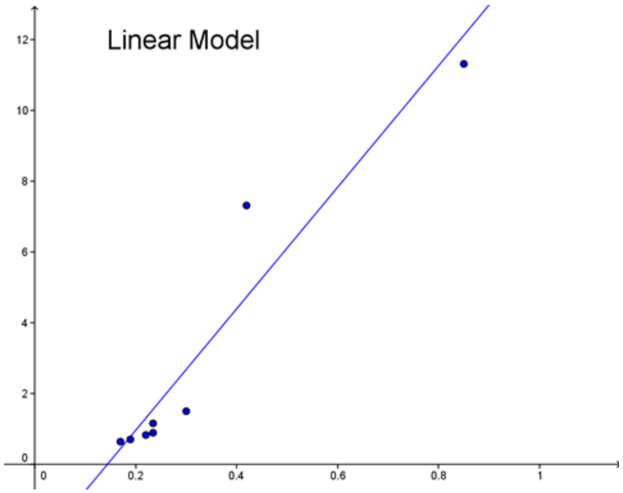
Time	110	120
Total Mass	913,254	880,377

Resource: www.nasa.gov

Prior Knowledge Needed:
Many of these data sets are available on the Data and Story Library (http://lib.stat.cmu.edu/DASL/). In some cases, only a partial data set is included here. Students should have a basic working knowledge of using data from a table to create a scatterplot (most likely using technology). Students should also know how to use technology to create a regression function (linear, quadratic, cubic, quartic, exponential, etc.) It is also helpful if the technology used provides a value of the coefficient of determination (r^2). If the technology provides a value for r^2, it will be useful for students to know that higher r^2 values imply better fits between the model and the data.
Common Core State Standards for Mathematical Content that support this example
Modeling is best interpreted not as a collection of isolated topics but rather in relation to other standards. Making mathematical models is a Standard for Mathematical Practice. Interpret functions that arise in applications in terms of the context Analyze functions using different representations Build a function that models a relationship between two quantities Construct and compare linear, quadratic, and exponential models and solve problems Make inferences and justify conclusions from sample surveys, experiments, and observational studies
Solutions
A complete solution is provided for Data Set 1 as an example. For other data sets, the data plot and the recommended model(s) are provided.

Data Set 1:

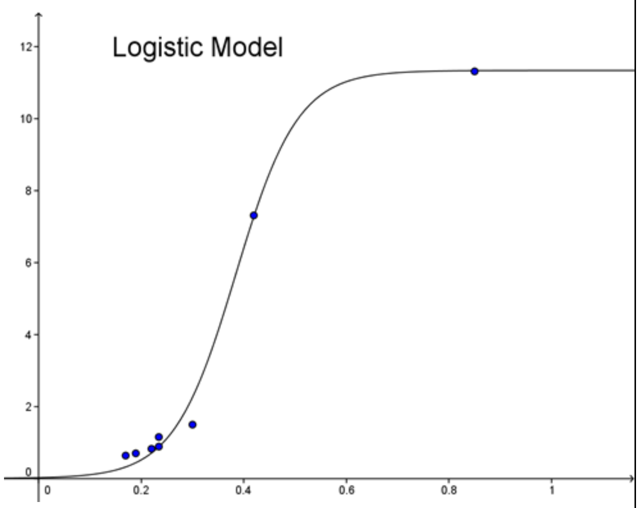
a) Based on the graph, the logistic model is the most likely best-fitting model. A linear model can be used as well. (Values are rounded to the nearest hundredth.)



Linear model: $G(d) = 17.16d - 2.48$, where d = diameter of the granules of sand in millimeters and G(d) represents the gradient of the slope of the beach in degrees. (Here, the correlation coefficient r is approximately 0.95, indicating that the model is a good fit for the data.)

Logistic model: $G(d) = \dfrac{11.33}{1 + 577.98e^{-16.54d}}$, where d = diameter of the granules of sand in millimeters and G(d) represents the gradient of the slope of the beach in degrees.

b) Graphs: The graphs of both models appear to the right.

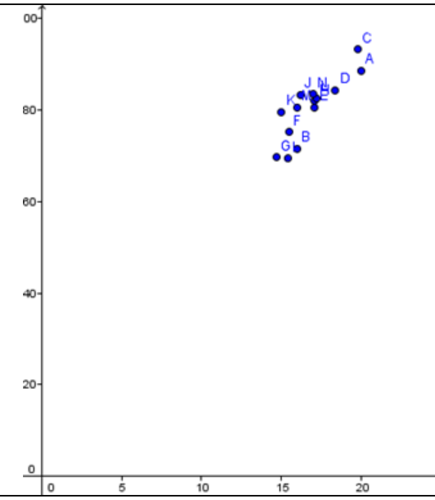


c) Limitations of the model: Both models can be used to predict values of the gradient based on values of the diameter of the granules of sand that lie between 0.17 mm (the smallest diameter given) and 0.85 mm (the largest diameter given). The models should not be used to predict values outside of this domain. Also note that the size of the granules of sand should not increase without bound; there should be a limiting value of the size of these granules. Similarly, there should be a limiting value of the gradient of the slope of the beach.

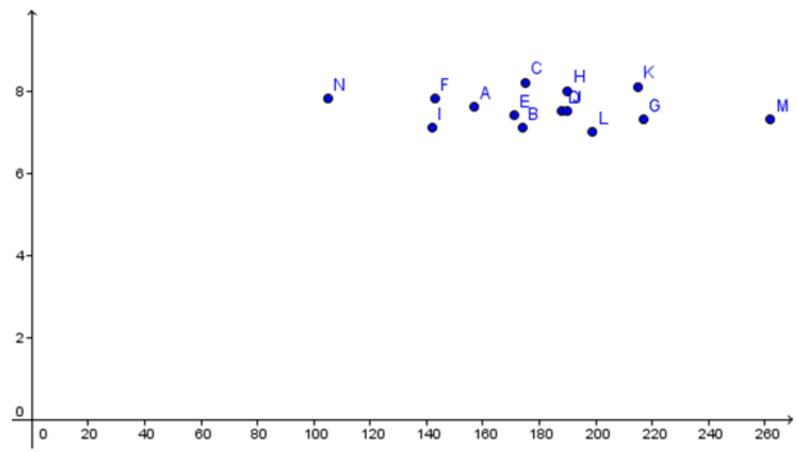
d) Sample Questions (other questions may be asked):

- For the linear model, what does the slope of the line represent? What does the y-intercept represent?
- Predict the gradient of the slope of the beach given a particular diameter not listed in the table.
- Predict the diameter of a granule of sand given a particular gradient of the slope of a beach not listed in the table.
- For the logistic model, what does the model suggest will happen to the gradient of the slope of a beach as the size of the granules of sand increases?

Data Set 2: This is a fairly common example of an application of linear regression. However, the data given can also be modeled using polynomials with approximately the same coefficient of determination as the linear model.

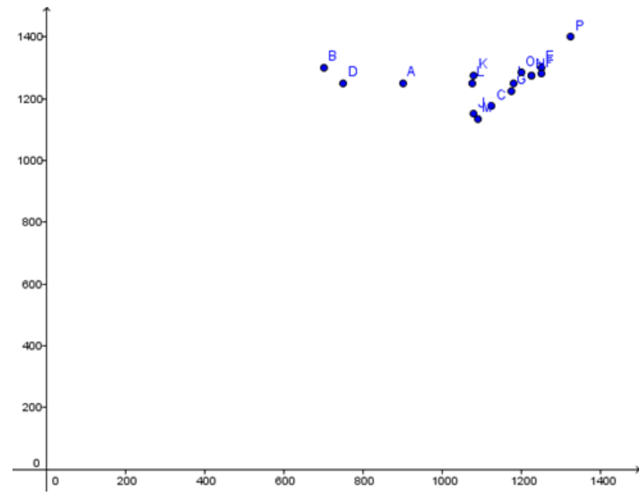


Data Set 3: This data is very scattered and does not fit the standard models very well. The data set is included for use as an example to illustrate that not all data sets can be modeled effectively.

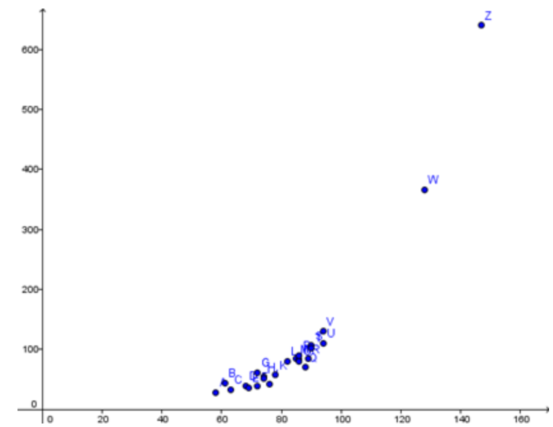


Data Set 4: This data set is not as scattered as data set 3, but there seems to be two "trends" in the graph of the data. Therefore, the standard models do not provide a close fit. A comparison of the models indicates that a cubic or quartic model fits most closely, with a quadratic model only a little less effective.

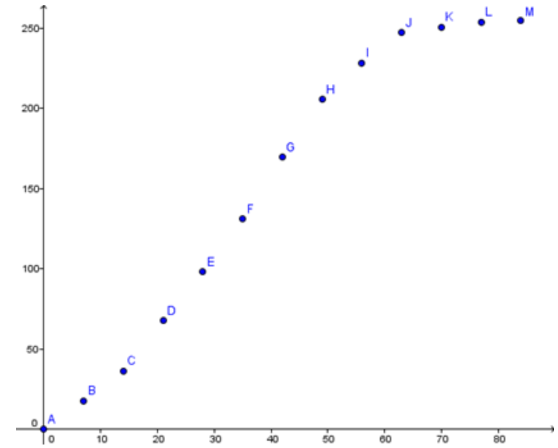
One way to improve the modeling process is to consider not the "initial year-ending year" data but to consider "initial year-*estimated length of occupation*" data that can be developed by finding the difference in the ending year and the initial year for each pueblo. This data gives rise to better-fitting models, with quadratic, cubic, and quartic models providing the best fits. Logarithmic and linear models also provide good fits, but these are not as close as the polynomial models.



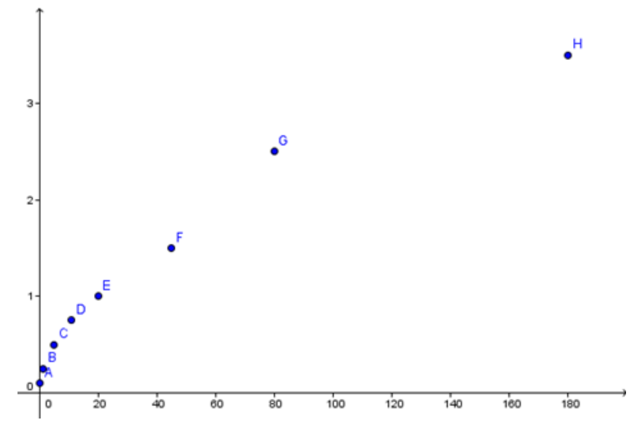
Data Set 5: An exponential model provides the best fit, followed closely by a quadratic model.



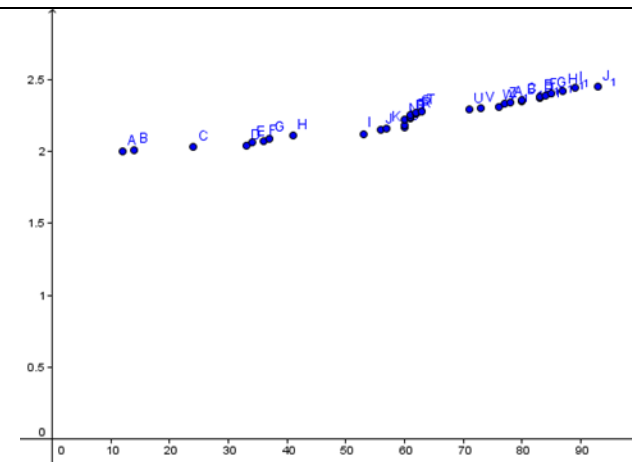
Data Set 6: The data illustrate a classic logistic model. However, students may have difficulty with the domain in trying to find the logistic model. The basic logistic model has a horizontal asymptote at y = 0. Since the data provided includes the point (0, 0) in the graph, including this point in the logistic calculation will result in a domain error.



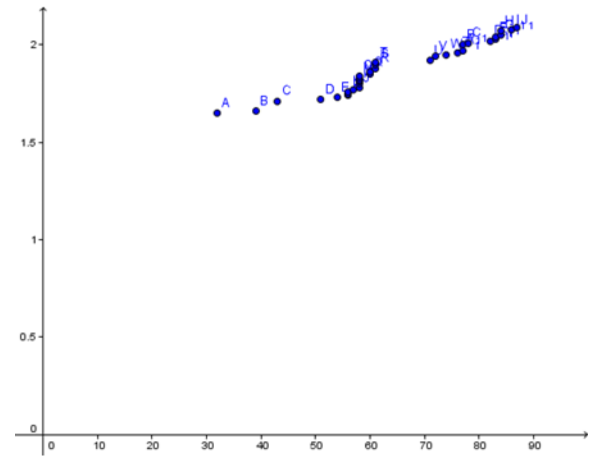
Data Set 7: A quadratic model provides the best fit. However, as with Data Set 1, there should be limits on the size of the particles being carried by the river.



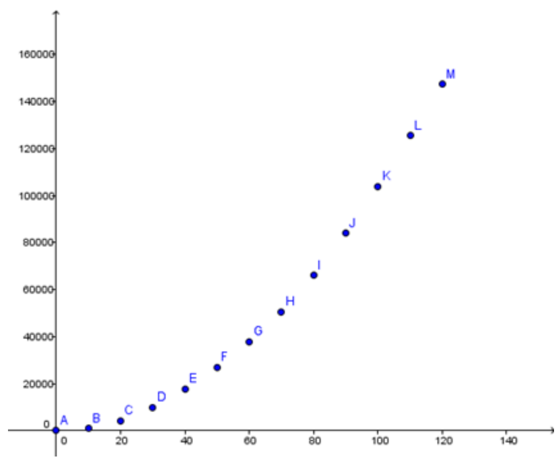
Data Set 8: A linear model provides the best fit. (Note: The given graph uses the number of years after 1900 as the x-value.)



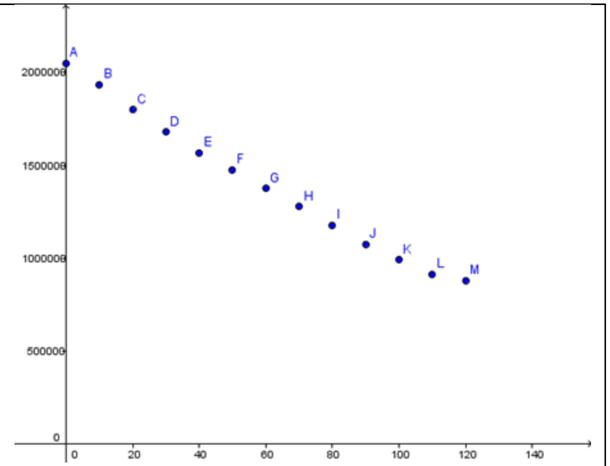
Data Set 9: A linear model provides the best fit. (Note: The given graph uses the number of years after 1900 as the x-value.)



Data Set 10: A quadratic model provides the best fit. (Note: This data is included as part of an educational project provided by NASA.)



Data Set 11: A quadratic model provides the best fit. (Note: This data is included as part of an educational project provided by NASA.)



Data Set 1: Sand Granules

For naturally occurring ocean beaches, the data below were collected. Find a function to model the relationship between the median diameter of granules of sand (in millimeters) and the gradient of the slope of the beach (in degrees).

Diameter	0.17	0.19	0.22	0.235	0.235	0.3	0.35	0.42	0.85
Gradient	0.63	0.7	0.82	0.88	1.15	1.5	4.4	7.3	11.3

a) Determine what type of model is most appropriate for the data. Explain what models were considered and why one model was chosen over the others. Explain what each of the variables represents.
b) Present graphical evidence that the model is valid for the data presented.
c) Explain any limitations of the model.
d) Create several questions that can be answered using the model. Answer the questions that are created.

Resource: *Physical Geography*, by A. M. King, Oxford Press, England

Data Set 2: Cricket Chirps

The following data relate the number of chirps per second of the striped ground cricket and the temperature in degrees Fahrenheit. Find a function to model the relationship between these two quantities.

a) Determine what type of model is most appropriate for the data. Explain what models were considered and why one model was chosen over the others. Explain what each of the variables represents.
b) Present graphical evidence that the model is valid for the data presented.
c) Explain any limitations of the model.
d) Create several questions that can be answered using the model. Answer the questions that are created.

Chirps/sec	20.0	16.0	19.8	18.4	17.1	15.5	14.7
Temperature	88.6	71.6	93.3	84.3	80.6	75.2	69.7

(table continued)

Chirps/sec	17.1	15.4	16.2	15.0	17.2	16.0	17.0
Temperature	82	69.4	83.3	79.6	82.6	80.6	83.5

Resource: *The Song of Insects*, by Dr. G. W. Pierce, Harvard College Press

Data Set 3: Ground Water Survey

The following data were collected from a random sample of wells in Northwest Texas. The data relate the bicarbonate (in parts per million) of the well water with the pH of the well water. Use the data to develop a model for the relationship between these two quantities.

a) Determine what type of model is most appropriate for the data. Explain what models were considered and why one model was chosen over the others. Explain what each of the variables represents.
b) Present graphical evidence that the model is valid for the data presented.
c) Explain any limitations of the model.
d) Create several questions that can be answered using the model. Answer the questions that are created.

Bicarbonate	157	174	175	188	171	143	217	190	142	190	215	199	262	105
pH	7.6	7.1	8.2	7.5	7.4	7.8	7.3	8.0	7.1	7.5	8.1	7.0	7.3	7.8

Resource: Union Carbide Technical Report K/UR-1

Data Set 4: Prehistoric Pueblos

The data below relate the estimated year of initial occupation with the estimated year of the end of occupation of prehistoric pueblos in a random sample of such pueblos in Utah, Arizona, and Nevada. Use the data to develop a model for the relationship between these two quantities.

a) Determine what type of model is most appropriate for the data. Explain what models were considered and why one model was chosen over the others. Explain what each of the variables represents.
b) Present graphical evidence that the model is valid for the data presented.
c) Explain any limitations of the model.
d) Create several questions that can be answered using the model. Answer the questions that are created.

Initial Year	900	700	1125	750	1250	1250	1175	1225
End Year	1250	1300	1175	1250	1300	1280	1225	1275

(table continued)

Initial Year	1180	1080	1080	1075	1090	1225	1200	1325
End Year	1250	1150	1275	1250	1135	1275	1285	1400

Resource: *Prehistoric Pueblo World*, by A. Adler, University of Arizona Press

Data Set 5: The Size of Alligators

Many wildlife populations are monitored by taking aerial photographs. For example, the length of an alligator can be estimated quite accurately from an aerial photograph. However, the alligator's weight is much more difficult to determine. The data below show the length (in inches) and weight (in pounds) of alligators captured in central Florida. Use the data to develop a model for the relationship between the length and weight of alligators in central Florida.

a) Determine what type of model is most appropriate for the data. Explain what models were considered and why one model was chosen over the others. Explain what each of the variables represents.
b) Present graphical evidence that the model is valid for the data presented.
c) Explain any limitations of the model.
d) Create several questions that can be answered using the model. Answer the questions that are created.

Length	58	61	63	68	69	72	72	74	74	76	78	82	85	86
Weight	28	44	33	39	36	38	61	54	51	42	57	80	84	83

(table continued)

Length	86	86	88	89	90	90	94	94	114	128	147
Weight	80	90	70	84	106	102	110	130	197	366	640

Resource: Exploring Data website, http://curriculum.qed.qld.gov.au/kla/eda, Education Queensland, 1997

Data Set 6: Sunflower Height

The data below relates the height of single sunflower plant (in cm) to the number of days the plant has been growing. Use the data to develop a model for the relationship between the height and the number of days the plant has been growing.

a) Determine what type of model is most appropriate for the data. Explain what models were considered and why one model was chosen over the others. Explain what each of the variables represents.
b) Present graphical evidence that the model is valid for the data presented.
c) Explain any limitations of the model.
d) Create several questions that can be answered using the model. Answer the questions that are created.

Day	0	7	14	21	28	35	42
Height	0.00	17.93	36.36	67.76	98.10	131.00	169.50

(table cont.)

Day	49	56	63	70	77	84
Height	205.50	228.30	247.10	250.50	253.80	254.50

Resource: Reed, H.S. and Holland, R.H. (1919), Growth of sunflower seeds; Proceedings of the National Academy of Sciences, volume 5, p. 140.

Data Set 7: Moving River Particles

Rivers and streams carry small solid particles of rock and mineral downhill, either suspended in the water column or bounced, rolled, or slid along the river bed. The data below show the speed (in m/sec) necessary to carry particles in suspension as it relates to the diameter of the particle (in mm). Use the data to develop a model for the relationship between the size of the particle in suspension and the speed necessary to move the particle.

a) Determine what type of model is most appropriate for the data. Explain what models were considered and why one model was chosen over the others. Explain what each of the variables represents.
b) Present graphical evidence that the model is valid for the data presented.
c) Explain any limitations of the model.
d) Create several questions that can be answered using the model. Answer the questions that are created.

Diameter	0.2	1.3	5	11	20	45	80	180
Speed	0.10	0.25	0.50	0.75	1.00	1.50	2.50	3.50

Resource: Nielsen, A. (1950) *Oikos,* 2, 176-96 as reported in Ecology for Environmental Sciences, Anderson J.M.

Data Set 8: World Records for Men's High Jump

The data below show the world record for the men's high jump (in meters) and the year in which the world record was set (between 1912 and 1993). Use the data to create a model for the relationship between the world record for the men's high jump and the year in which the record was set.

a) Determine what type of model is most appropriate for the data. Explain what models were considered and why one model was chosen over the others. Explain what each of the variables represents.
b) Present graphical evidence that the model is valid for the data presented.
c) Explain any limitations of the model.
d) Create several questions that can be answered using the model. Answer the questions that are created.

Year	1912	1914	1924	1933	1934	1936	1937	1941	1953	1956	1957	1960	1960
Record	2.00	2.01	2.03	2.04	2.06	2.07	2.09	2.11	2.12	2.15	2.16	2.17	2.18

(table continued)

Year	1960	1961	1961	1961	1962	1962	1963	1971	1973	1976	1976	1977	1978
Record	2.22	2.23	2.24	2.25	2.26	2.27	2.28	2.29	2.30	2.31	2.32	2.33	2.34

(table continued)

Year	1980	1980	1983	1983	1984	1985	1985	1987	1989	1993
Record	2.35	2.36	2.37	2.38	2.39	2.40	2.41	2.42	2.44	2.45

Data Set 9: World Records for Women's High Jump

The data below show the world record for the women's high jump (in meters) and the year in which the world record was set (between 1932 and 1987). Use the data to create a model for the relationship between the world record for the women's high jump and the year in which the record was set.

a) Determine what type of model is most appropriate for the data. Explain what models were considered and why one model was chosen over the others. Explain what each of the variables represents.
b) Present graphical evidence that the model is valid for the data presented.
c) Explain any limitations of the model.
d) Create several questions that can be answered using the model. Answer the questions that are created.

Year	1932	1939	1943	1951	1954	1956	1956	1956	1957	1958	1958	1958	1958
Record	1.65	1.66	1.71	1.72	1.73	1.74	1.75	1.76	1.77	1.78	1.80	1.81	1.82

(table continued)

Year	1958	1958	1960	1960	1961	1961	1961	1971	1972	1974	1976	1977	1977
Record	1.83	1.84	1.85	1.86	1.88	1.90	1.91	1.92	1.94	1.95	1.96	1.97	2.00

(table continued)

Year	1978	1982	1983	1983	1984	1984	1986	1987
Record	2.01	2.02	2.03	2.04	2.05	2.07	2.08	2.09

Data Set 10: Space Shuttle Ascent Altitude

On July 4, 2006, the Space Shuttle Discovery launched from Kennedy Space Center on mission STS-121 to begin a rendezvous with the International Space Station. The data below show the altitude (in feet) of Discovery every 10 seconds from liftoff to the separation of the solid rocket boosters. Use the data to model the altitude of Discovery at time t.

a) Determine what type of model is most appropriate for the data. Explain what models were considered and why one model was chosen over the others. Explain what each of the variables represents.
b) Present graphical evidence that the model is valid for the data presented.
c) Explain any limitations of the model.
d) Create several questions that can be answered using the model. Answer the questions that are created.

Time	0	10	20	30	40	50	60	70
Altitude	7	938	4160	9872	17635	26969	37746	50548

(table cont.)

Time	80	90	100	110	120
Altitude	66033	83966	103911	125512	147411

Resource: www.nasa.gov

Data Set 11: Space Shuttle Ascent Total Mass

On July 4, 2006, the Space Shuttle Discovery launched from Kennedy Space Center on mission STS-121 to begin a rendezvous with the International Space Station. The data below show the total mass (in kg) of Discovery every 10 seconds from liftoff to the separation of the solid rocket boosters. Use the data to model the total mass of Discovery at time t.

a) Determine what type of model is most appropriate for the data. Explain what models were considered and why one model was chosen over the others. Explain what each of the variables represents.
b) Present graphical evidence that the model is valid for the data presented.
c) Explain any limitations of the model.
d) Create several questions that can be answered using the model. Answer the questions that are created.

Time	0	10	20	30	40	50
Total Mass	2,051,113	1,935,155	1,799,290	1,681,120	1,567,611	1,475,282

(table continued)

Time	60	70	80	90	100	110	120
Total Mass	1,376,301	1,277,921	1,177,704	1,075,683	991,872	913,254	880,377

Resource: www.nasa.gov